1 Motivation

Consider a very simple framework for analysing the causal effect of a single unit in the treatment unit $i = 0$ and two units in the control group $i = 1, 2$. It is assumed that before the intervention at time period $t = T_0$ the units have a joint distribution of the form[1]

$$\boldsymbol{y} = \begin{pmatrix} Y_0 \\ Y_1 \\ Y_2 \end{pmatrix} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \quad \text{before } T_0$$

where $\boldsymbol{\mu} = (\mu_0, \mu_1, \mu_2)'$ and $\boldsymbol{\Sigma}$ is some positive definite covariance matrix with Choleski decomposition $\boldsymbol{\Sigma} = \boldsymbol{R}\boldsymbol{R}'$ and $\boldsymbol{R}$ is a *upper* triangular matrix. Assume that the intervention affects the mean of the first variable such that $\mathbb{E}(Y_0) = \mu_0 + \delta$ after the intervention, whereas the means of the other variables remain unaffected. Accordingly δ represents the treatment effects on Y_0.

We are interested to derive an optimal estimator for the counterfactual

$$\widehat{Y}_0^N = \mathbb{E}(Y_0 | \delta = 0, Y_1, Y_2) \quad \text{after } T_0$$

Let $Q = R^{-1}$ and q denotes the first row of Q, then

$$q'y = q'\mu + \varepsilon$$

where $\varepsilon \sim \mathcal{N}(0, 1)$ with $\mathbb{E}(\varepsilon | Y_1, Y_2) = 0$. It follows that

$$\begin{aligned} \widehat{Y}_0^N &= w_1 Y_1 + w_2 Y_2 + \mu^* \\ &= \mu_0 + w_1(Y_1 - \mu_1) + w_2(Y_2 - \mu_2) \end{aligned}$$

where $w_1 = -q_2/q_1$, $w_2 = -q_3/q_1$ and $\mu^* = \mu_0 - w_1\mu_1 - w_2\mu_2$. These results imply that there is no reason to impose the restrictions $w_1 \geq 0$, $w_2 \geq 0$ (positivity) and $w_1 + w_2 = 1$ (adding-up restriction). Furthermore, the construction of the synthetic control should include a constant term, as otherwise the synthetic control may have a different mean. See also Doudchenko and Imbens (2017) for

[1] For the ease of exposition we suppress the time index t as in this section we neglect any dynamic effects which will be considered in the next section.

a careful discussion of these restriction.

For illustration assume that

$$y \sim \mathcal{N}\left(\begin{pmatrix}1\\1\\1\end{pmatrix}, \begin{pmatrix}1 & 0.1 & 0.4\\0.1 & 1 & 0.5\\0.4 & 0.5 & 1\end{pmatrix}\right)$$

For this example the optimal weights for the synthetic control result as $w_1 = -0.133$, $w_2 = 0.4667$ and $\mu^* = 1 - w_1 - w_2 = 0.667$. Note that w_1 is negative even if all bivariate correlations between the individuals are positive. One may argue that this solution does not make much sense as it is not clear what is means that Y_1 enters the synthetic control with a negative sign. This demonstrates the trade-off between the optimality in a statistical sense and the economic interpretability of the solution.

What happens if we impose the restrictions that all weights are positive and sum up to unity? In this case the restricted optimum yields the linear combination $\widetilde{Y}_0^N = 0.2Y_1 + 0.8Y_1$. Since all units have the same mean, the restricted solution is unbiased, that is, $\mathbb{E}(\widetilde{Y}_0^N) = \mu_0$. The important difference is in the variance of these estimates. For our example we obtain

$$var(Y_0 - \widehat{Y}_0^N) = 0.827$$
$$var(Y_0 - \widetilde{Y}_0^N) = 1.16$$

It is interesting to note that the variance of the restricted estimate is even larger than the unconditional variance of Y_0. This is possible as $(w_1, w_2) = (0, 0)$ is not included in the restricted parameter space.

It is not difficult to see that if Y_0 is not correlated with Y_1 and Y_2, then the optimal estimate boils down to $\widehat{Y}_0^N = \mu_0$ and, therefore, it does not make sense to involve a synthetic control. In microeconometric studies it is usually assumed that the individuals in the treatment group and the individuals in the control group are uncorrelated. In such cases we do not care about constructing a synthetic control. The crucial feature of synthetic control methods is the correlation between the units in the treatment and control group. In macroeconomic applications the variables in the treatment and control groups are typically correlated and it is therefore important to model the relationship between the variables.

In empirical practice it is often the case that the number of pre-intervention time periods T_0 is small and may even be smaller than k, the number of units in the control group. In this cases some kind of regularization is necessary to obtain a reliable estimate of $\widehat{Y}_0^N$. Doudchenko and Imbens (2017) suggest to invoke the elastic net penalty when estimating the weights. Instead of just shrinking the parameters towards zero we adopt a penalty that is flexible enough to produce more reasonable weighting schemes by using the objective function

$$Q(w, \lambda_1, \lambda_2) = \sum_{t=1}^{T_0} \left(y_{0t} - \mu^* - \sum_{i=1}^{k} w_i y_{it} \right)^2 + \lambda_1 \left(\sum_{i=1}^{k} w_i^2 \right) + \lambda_2 \left(1 - \sum_{i=1}^{k} w_i \right)^2$$

The first part of the penalty weighted by the shrinkage parameter λ_1 is the usual regularization penalty that shrinks the weights towards zero. The second part of the penalty forces the sum of the weights towards unity. Note that if $\lambda_1 \to \infty$ and $\lambda_2 \to \infty$, then the weights converge to equal weights $1/k$ which appears to yield a more reasonable target than shrinking merely towards zero. The estimator can easily be estimated as it has an explicit solution given by

$$\widehat{w}|_{\lambda_1,\lambda_2} = (X'X + \lambda_1 I_k + \lambda_2 \mathbf{1}_k \mathbf{1}_k')^{-1} \left(X'y^0 + \lambda_2 \mathbf{1}_k \right).$$

where X is a $T_0 \times k$ matrix that stacks all observations for $t = 1, \ldots, T_0$ and $i = 1, \ldots, k$ and y^0 is a $T_0 \times 1$ vector stacking the T_0 time series observations of Y_0. In practice the shrinkage parameters can be chosen by cross validation, where our experience suggest that optimising subject to the restriction $\lambda_1 = \lambda_2$ reduces the computing time and already produces reasonable estimates.

2 Dynamic models

When modelling macroeconomic time series it is often assumed that the $(k+1) \times 1$ vector of time series $y_t = (Y_{0t}, \ldots, Y_{kt})'$ can be represented by a vector autoregressive model given by

$$\begin{aligned} y_t &= \alpha + A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t \\ y_t &= \mu + A(L)(y_{t-1} - \mu) + u_t \end{aligned}$$

where $\mu = (\mu_0, \mu_1, \ldots, y_k)'$, $\mathbb{E}(u_t u_t') = \Sigma$ is a positive definite covariance matrix. In this section we assume that the vector of time series can be represented by such a VAR process up to the intervention period T_0. To analyse the effect of an intervention on the first time series we consider two scenarios. First we assume that the intervention changes the unconditional mean of the first series μ_0 to $\mu_0 + \delta$, where δ represents the 'treatment effect'.[2] The second scenario is more common in empirical macroeconomics, where the intervention is represented by a 'structural shock'.

Let us first investigate a change in the unconditional mean, where all other parameters remain the same. Using again the factorization $QQ' = \Sigma^{-1}$, where Q is upper triangular we have

$$q'y_t = q'\mu + q'A(L)(y_{t-1} - \mu) + q'u_t$$
$$y_{0t} = \mu_0 + \sum_{i=1}^{k} w_i(y_{it} - \mu_i) + \beta(L)'(y_{t-1} - \mu) + \epsilon_t$$

where $q' = (q_0, \ldots, q_k)$ denotes the first row of Q, $w_i = -q_{i+1}/q_1$, $\beta(L)' = q' + q'A_1 + \ldots + q'A_p$ and $\epsilon_t = q'u_t$. This representation implies that the counterfactual that we can "predict" y_{1t} by a linear combination of current and lagged values

$$\widehat{y}_t^N = \mu_0 + \sum_{i=1}^{k} w_i(y_{it} - \mu_i) + \beta(L)'(y_{t-1} - \mu) \quad \text{for } t = T_0 + 1, \ldots, T. \tag{1}$$

This result implies that the lags of the variable can only be ignored if $\beta(L) = 0$ which in turn requires that q is orthogonal to all columns of $A_1, \ldots, A_q$. Of course this highly unlikely to be the case in empirical practice.

An important problem with the optimal solution (1) is that it involves $(p + 1)(k + 1)$ parameters which may be difficult to estimate reliably in practice. We therefore focus on the static representation given by

$$y_{0t} = \mu^* + \sum_{i=1}^{k} w_i y_{it} + e_t \quad \text{for } t = 1, \ldots, T_0. \tag{2}$$

where $\mu^* = \mu_0 - \sum_{i=1}^{k} w_i \mu_i$. Obviously, neglecting the lags $\beta(L)'y_{t-1}$ results in an

[2]We also study the related situation where the first element of the vector α switches to a new value at T_0.

autocorrelated error e_t in general. To account for this autocorrelation we apply a two step approach for estimating the counterfactual $\widehat{y}_t^N$. In the first step the representation (2) is estimated as proposed in the previous section. The residual of the fitted equation ($\widehat{e}_t$) provides us with an estimate of e_t. For the residuals we fit the autoregression

$$\widehat{e}_t = \phi_1 \widehat{e}_{t-1} + \cdots + \phi_q \widehat{e}_{t-q} + \nu_t \tag{3}$$

and the counterfactual series for $t = T_0 + 1, \ldots, T$ is obtained as

$$\widehat{y}_t^N = \widehat{\mu}^* + \sum_{i=1}^{k} \widehat{w}_i y_{it} + \widehat{\phi}_1 \widehat{e}_{t-1} + \cdots + \widehat{\phi}_q \widehat{e}_{t-q} \ , \tag{4}$$

where $\widehat{\mu}^*$ and $\widehat{w}_i$ denotes the estimates of the static representation (2) and $\widehat{\phi}_j$ denote the least-squares estimates from (3).

Obviously, this two-step approach is inefficient as it imposes (common factor) restrictions to the dynamics. If sufficient data is available it is therefore desirable to estimate the dynamic equation (1) by including all necessary lags of the variables. In this case it may be necessary to impose L_1 or L_2 penalty terms to the least-squares objective function.